

Grif et les index electroniques

Hélène Richy

► To cite this version:

Hélène Richy. Grif et les index electroniques. [Rapport de recherche] RR-1756, INRIA. 1992. inria-00076996

HAL Id: inria-00076996

<https://hal.inria.fr/inria-00076996>

Submitted on 29 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNITÉ DE RECHERCHE
INRIA-RENNES

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél. (1) 39 63 55 11

Rapports de Recherche

1 9 9 2



ème

anniversaire

N° 1756

Programme 3

*Intelligence artificielle, Systèmes cognitifs et
Interaction homme-machine*

GRIF ET LES INDEX ÉLECTRONIQUES

Hélène RICHY

Octobre 1992



★ R R - 1 7 5 6 ★

Campus Universitaire de Beaulieu
35042 - RENNES CÉDEX
FRANCE
Téléphone : 99 84 71 00
Télex : UNIRISA 950 473F
Télécopie: 99 38 38,32

Grif et les index électroniques

Hélène RICHY

*Irisa, Campus universitaire de Beaulieu
F-35042 Rennes Cedex (France)
Tél: (33) 99 84 73 71 Fax: (33) 99 38 38 32
Adr. électronique: richy@irisa.fr*

Projet OPÉRA

Outils pour les documents électroniques, recherche et applications

Programme INRIA: 3

Intelligence artificielle, systèmes cognitifs et interaction homme-machine

Résumé: Ce rapport présente les index électroniques de Grif. Ces index sont à la fois des tables alphabétiques, dans lesquelles sont cités les termes importants d'un document, et des liens qui permettent de parcourir le document comme un hypertexte, sans utiliser la structure générique initiale du document.

Plus précis que les index imprimés, qui font uniquement référence aux pages d'un document, ces index permettent de retrouver directement dans le document électronique le passage, la phrase ou la section qui traitent du sujet cherché. De plus, ces index peuvent être utilisés aussi bien lors de la consultation d'un document qu'au cours de son édition.

Mots-clés: document structuré, hypertexte, index, Grif.

Electronic Index in Grif

Abstract: This report presents the electronic index available in the Grif system. This index is both an alphabetic list in which all the information of primary importance is described, and a set of links given for navigation within the document.

This electronic index can be used in a natural way as it is displayed in the same form as in a paper document. However, it is more precise, because it does not only indicate a page number, but it also shows the related passage within the structured document: this passage may be either a sentence, a few paragraphs or a whole section on the selected subject. Furthermore, this electronic index is available not only while browsing an electronic document, but also while editing an uncompleted document.

Keywords: structured document, hypertext, index, Grif.

Publication interne Irisa - n°677 Septembre 1992 -40 pages
Rapport de recherche INRIA 1756 Septembre 1992

Table des matières

1 Des index traditionnels aux index électroniques	3
1.1 Index de documents traditionnels	3
1.2 Index de documents électroniques	5
1.3 Index électronique	7
2 Les index électroniques dans Grif	9
2.1 Principe général de l'intégration dans Grif	9
2.2 Structure des index électroniques	12
2.3 Présentation des index électroniques	20
2.4 Construction des tables d'index	20
3 Interface utilisateur des index de Grif	22
3.1 Chargement des extensions	22
3.2 Vues	23
3.3 Marques	24
3.4 Descripteurs de marque	25
3.5 Descripteurs de renvoi	26
3.6 Options de construction	27
3.7 Construction des tables d'index	28
3.8 Liens électroniques	29
4 Résultats et perspectives	32
4.1 Premières expérimentations	32
4.2 Vers une indexation plus automatique	33
4.3 Vers une présentation plus homogène des termes	33
4.4 Conclusion et perspectives	33
4.5 Remerciements	34
5 Bibliographie	35

Grif et les index électroniques

Projet Opéra, Irisa, Rennes (France)

Résumé: Ce rapport présente les index électroniques de Grif. Ces index sont à la fois des tables alphabétiques, dans lesquelles sont cités les termes importants d'un document, et des liens qui permettent de parcourir le document comme un hypertexte, sans utiliser la structure générique initiale du document.

Plus précis que les index imprimés, qui font uniquement référence aux pages d'un document, ces index permettent de retrouver directement dans le document électronique le passage, la phrase ou la section qui traitent du sujet cherché. De plus, ces index peuvent être utilisés aussi bien lors de la consultation d'un document qu'au cours de son édition.

Mots-clés: document structuré, hypertexte, index, Grif.

Les outils de PAO permettent, avec un certain confort d'utilisation et un haut degré d'interactivité, de produire des documents dont la qualité typographique finale peut être excellente. Cependant, certaines tâches demandent encore beaucoup d'efforts aux auteurs, notamment la construction des index. Les interfaces proposées sont souvent complexes et les effets ne sont pas visibles de façon WYSIWYG. Cette tâche est rendue encore plus difficile puisque, généralement, elle concerne des documents volumineux dans lesquels les termes à indexer sont dispersés.

Dans ce rapport, nous nous intéressons à la production de tables d'index dans un document structuré et supposons que l'auteur est capable de choisir les « bons » termes à indexer, en attendant que les suggestions d'indexation proposées par les méthodes d'indexation automatique [Salton 89a] [Salton 89b] puissent aider l'auteur dans ses choix.

La réalisation que nous décrivons ici résulte de la convergence de deux approches, celle des *hypertextes* et celle des *documents structurés*. L'objectif est de faciliter la production des tables d'index dans des documents structurés: les index sont construits comme des liens non-hiérarchiques (comme dans un hypertexte) sur la structure logique hiérarchique des documents (comme dans un document structuré).

En s'appuyant sur la structure générique initiale, l'auteur peut « marquer » n'importe quel élément d'un document, aussi bien un groupe de mots qu'un chapitre, une section ou une succession de paragraphes. À sa demande, une table d'index est créée dans le document à partir de ces marques. L'*index électronique* produit est aussi un document structuré comportant des liens d'hypertexte facilitant non seulement la consultation des documents (par le lecteur), mais aussi la saisie des index et la rédaction du document (par

l'auteur). Comme dans un hypertexte, il suffit de cliquer sur un terme pour faire afficher la partie du document concernée. Ces liens électroniques facilitent également la création des index : il suffit de cliquer dans le document sur un terme marqué pour l'indexation, pour faire afficher toutes les informations associées (clé choisie, glose, etc.) et pouvoir ainsi les éditer.

Ce rapport est composé de quatre parties. La première présente la notion d'index, depuis les index traditionnels jusqu'aux index de documents électroniques. La seconde présente la mise en œuvre des *index électroniques* dans l'éditeur structuré Grif. La troisième décrit son interface utilisateur. Enfin, la dernière partie analyse les premiers résultats obtenus.

1 Des index traditionnels aux index électroniques

1.1 Index de documents traditionnels

Les index sont des tables alphabétiques dans lesquelles sont cités tous les noms et sujets traités, importants dans le cadre de l'ouvrage concerné. Ces termes sont accompagnés des références aux pages où se trouvent les textes qui s'y rapportent. Ils peuvent être précisés par une glose. Par exemple, dans l'entrée:

Sⁱ Jacques (Aéroport de Rennes): 10, 15, voir aussi Rennes

Sⁱ Jacques est le terme, *Aéroport de Rennes* est une glose, *10, 15* sont les numéros des pages qui décrivent cet aéroport et *voir aussi Rennes* est un renvoi à l'entrée *Rennes* de l'index.

Les index sont essentiels pour faciliter la consultation de documents volumineux, tels que les notices techniques ou scientifiques. Autrefois, la construction d'un index se faisait à partir d'un manuscrit, dont les feuillets étaient préalablement numérotés et dans lequel l'auteur avait souligné les mots ou groupes de mots importants [Dufour 71]. Chaque mot ou groupe de mots donnait ensuite lieu à l'établissement d'une fiche sur laquelle était reporté le numéro de la page qui le mentionne. Enfin, ces fiches étaient classées manuellement par ordre alphabétique et par ordre croissant des numéros de pages. Elles constituaient alors le manuscrit de l'index de l'ouvrage.

Le plus grand soin devait être apporté à la constitution de ces fiches afin de supprimer les doubles (mêmes termes, mêmes numéros de page), de subdiviser les entrées contenant un trop grand nombre de références, d'ajouter des fiches de renvoi pour éviter des répétitions inutiles et faciliter les recherches du lecteur, de vérifier l'orthographe des termes, etc.

Le savoir-faire en ce domaine permet de formuler un certain nombre de recommandations, aussi bien pour le choix des termes que pour leur présentation dans une table d'index. Il est utile de les rappeler ici, dans la mesure où certaines d'entre elles sont encore totalement ignorées des systèmes de PAO actuels.

1.1.1 Choix des termes

Les termes retenus sont ceux qui sont susceptibles d'être recherchés par le lecteur. Le choix des termes est généralement fait par l'auteur lui-même à l'intention des lecteurs de l'ouvrage. Les groupes de mots relevés dans l'ouvrage sont généralement ceux qui figureront comme entrée dans les index. Toutefois, des renvois peuvent permettre de choisir un terme mieux adapté aux recherches du lecteur.

Pour les groupes de mots, l'entrée doit indiquer en premier le mot le plus important par rapport au contexte. Par exemple, «structure de document» ou «document (structure de)» selon que c'est la structure ou le document qui revêt l'aspect le plus significatif, ou bien les deux à la fois. Des sous-entrées peuvent être définies à partir de la même entrée. Par exemple, «structure de _» et «présentation de _» sont deux sous-entrées de l'entrée «document». Pour les

noms de personne, le nom propre doit figurer le premier, suivi des initiales du (ou des) prénom(s) : par exemple, on écrira généralement *Balzac, H. de* et non *H. de Balzac* (figures 1 et 17).

On peut éventuellement distinguer différentes catégories d'entrées qui seront repérées dans l'index par une différence typographique : par exemple, les noms de lieu seront présentés en gras, les noms propres en italique, etc.

Les index peuvent aussi être construits selon un classement thématique : index des noms de personne, index des lieux géographiques, etc. Un ouvrage peut alors avoir plusieurs index.

Brauen, T.L., 324-325, 368
Bresnan, J., 396, 421
Briscoe, T., 382, 319
Briscoe, E.J., 388, 420
Buckley, C., 255, 273, 292, 310-311, 348, 361, 370

Figure 1: *Extrait d'un index d'auteurs*

1.1.2 Présentation des index

La présentation des index doit assurer à l'utilisateur une recherche aisée et rapide. Au besoin, une note introductive donne au lecteur les explications nécessaires sur la signification des éventuelles différences typographiques ou sur le mode de classement adopté, s'il y a plusieurs index.

((left parenthesis), 51, 134, <i>140</i> , <i>145-150</i> , 345.
\ (, 409.
) (right parenthesis), 51, 134, <i>140</i> , <i>145-150</i> , 345.
[(left bracket), 51, 134, <i>146-148</i> , 171, 408, 437.
[], <u>28</u> , <u>79</u> , 302.

Figure 2: *Extrait de l'index du T_EXbook [Knuth 84]*

En général les entrées sont triées par ordre alphabétique. Dans les manuels techniques, les entrées peuvent contenir des caractères non-alphabétiques. Ces entrées sont alors placées en début (ou en fin) d'index et triées dans un ordre quelconque (voir la figure 2).

Les références sont toujours triées par ordre croissant : ces références peuvent être des numéros de page, de section, de volume¹ ou encore de figure ou de note.

1. En ce sens, on peut considérer une bibliographie comme un index qui renvoie à d'autres documents ou volumes.

Une typographie différente (numéro de page en gras, par exemple) peut permettre de distinguer les renvois importants. L'application des règles typographiques permet de produire des index dont la lecture sera plus aisée. Ces règles concernent la ponctuation (entre les références aux pages, dans les intervalles de page, après une entrée ou une sous-entrée), l'usage des lettres majuscules et minuscules, la typographie des noms de personne, l'espacement des groupes de même initiale, le regroupement des entrées ayant des initiales peu usitées (J-K, X-Z), les mentions utilisées pour les renvois (*voir*, *voir aussi* ou *cf.* et *cf. aussi*), etc.

Un certain nombre de variations sont autorisées : usage ou non de ponctuation finale dans les entrées, usage de capitale initiale, autorisation des sous-entrées, présentation sur deux colonnes, etc. Quelles que soient les règles de présentation choisies, celles-ci doivent être respectées par **toutes** les entrées d'une table d'index afin de garantir une présentation homogène et une bonne lisibilité de la table (cf. figures 1, 2 et 3).

L
Le Breton (Auguste): 112
Li (symbole chimique du <i>lithium</i>): 73
Li (mesure chinoise): <i>voir</i> Mesures
Littérature (évolution de la): 107, 119-123 , <i>voir aussi</i> Théâtre
africaine: 12, 102, 125
américaine: 34, 104
européenne: chapitre 4, 106

Figure 3: Quelques entrées d'une table d'index

1.2 Index de documents électroniques

De gros efforts ont été faits pour faciliter la production d'index dans les documents électroniques [Aurbach 87] [Bentley 88] [Chen 88] [Lamport 87] [Louarn 90]. Effectivement, avec l'introduction de la PAO, les méthodes de travail ont changé : finis la numérotation manuelle des feuillets et l'établissement des fiches de classement alphabétique. La mise en page des documents est calculée par les outils de l'édition électronique. Le classement alphabétique des termes est automatisé. La plupart des éditeurs informatiques proposent des outils intégrés permettant la construction des tables d'index et leur présentation. Les mots concernés sont simplement «marqués» dans le texte pour extraction et classement. Cependant, malgré cette automatisation, un certain nombre de tâches fastidieuses restent encore à faire pour obtenir des tables d'index dont le contenu serait satisfaisant.

La plupart des traitements de texte demandent aux utilisateurs de procéder en deux temps : tout d'abord, insérer des marques d'indexation dans le document, ensuite demander la génération des tables d'index à partir des marques d'indexation insérées précédemment.

1.2.1 Pose des marques

Avant de créer un index, il faut choisir les éléments à indexer. Les termes choisis doivent être « indiqués » dans le document par l'utilisateur. Ce choix est effectué en insérant dans le document des *marques d'index* qui indiquent deux choses : d'une part, l'emplacement du texte dans le document, qui sera référencé par son numéro de page dans la table d'index et d'autre part, la clé qui sera utilisée comme entrée dans la table d'index pour désigner cette référence.

Pour les traitements de texte tels que Word [Word 89], FrameMaker [Frame-maker 90] ou encore pour un formateur comme LaTeX [Lamport 86], cet emplacement est ponctuel : une marque d'indexation est placée entre deux caractères quelconques. C'est la position de cette marque qui déterminera le numéro de la page qui figurera dans l'index.

Avec FrameMaker, les marqueurs peuvent avoir les contenus suivants :

Littérature, voir Théâtre <\$nopage>

pour définir un renvoi, sans numéro de page

Li (symbole du <eti-ital>lithium<Default Para Font>)

pour mettre un mot en italique dans une entrée

Littérature:africaine <eti-principale>

pour définir deux niveaux d'entrée et le format du numéro de page (gras).

Avec Microsoft Word4 (sur MacIntosh), il faut introduire dans le texte :

.i.Littérature (évolution de la)#(voir Théâtre);

pour définir un renvoi sans numéro de page

.iB.Littérature:africaine;

pour définir une sous-entrée avec le numéro de page en gras (B = bold).

Avec LaTeX, la commande MakeIndex permet d'inclure des sous-clés :

\index{Litt'érature!africaine}

pour définir les deux niveaux d'entrée.

Il est généralement possible d'étendre la portée d'une référence. Pour indiquer, non plus une page de référence, mais une suite de pages (un intervalle), il suffit de placer dans le document une seconde marque ponctuelle ayant la même clé que la première marque et d'ajouter une indication complémentaire précisant qu'il s'agit du début (première marque) ou de la fin (seconde marque) d'une zone de référence. Par exemple, pour FrameMaker, il faut ajouter \$startrange dans la première marque et \$endrange, dans la seconde marque. Avec Word, il suffit de placer une parenthèse ouvrante ou une parenthèse fermante pour grouper deux marques qui définiront un intervalle.

Généralement, la clé choisie est identique à un mot du texte. Avec Word et FrameMaker c'est, par défaut, le texte sélectionné qui est choisi comme clé. Mais il est toujours possible de modifier cette clé, soit pour en choisir une autre, soit pour utiliser plusieurs niveaux de clé ou pour définir des entrées multiples (plusieurs clés définies pour une seule marque).

Associé à cette marque, on peut trouver un certain nombre d'informations qui serviront à la construction de la ou des tables d'index: le format de l'entrée, les sous-entrées éventuelles, la catégorie, la typographie de la référence, un commentaire, etc.

1.2.2 Génération des tables d'index

À la demande de l'utilisateur, l'éditeur analyse toutes les marques, calcule les numéros des pages, trie par ordre alphabétique, fusionne les entrées doubles et supprime les duplicata de pages de référence. Puis, l'index obtenu est placé à la fin du document, après une marque de division (avec Word) ou dans un document séparé (avec FrameMaker).

Un certain nombre d'options peuvent être choisies pour modifier la structure, la présentation ou le contenu de ces tables d'index. Selon le traitement de texte utilisé, ces options sont plus ou moins nombreuses. Avec Word, par exemple, il est possible, pour obtenir un résultat plus rapide de ne demander la construction que d'une partie de l'index.

Les règles de présentation sont gérées par ces éditeurs de la même manière que dans les autres parties du document. Cette approche présente certains avantages, en particulier la possibilité de modifier facilement les entrées (les styles des paragraphes de Word ou les formats de paragraphes de FrameMaker). Par contre, il n'y a aucun contrôle sur le contenu des index:

- les renvois sont considérés comme des commentaires;
- les références indiquées dans les renvois peuvent ne pas exister dans la table d'index;
- aucun format n'est imposé pour les renvois; il n'y a donc aucune garantie d'homogénéité de présentation de ces renvois: par exemple, on peut trouver des renvois introduits par *cf.* ou par *voir* dans une même table.

1.2.3 Mise au point des tables

Une fois ces deux premières étapes effectuées, le résultat est rarement satisfaisant, aussi bien pour le contenu que pour la présentation des index obtenus: tables incomplètes ou trop denses, références mal réparties dans le document, etc.

La table d'index générée permet d'avoir une vision globale des marques d'indexation qui ont été introduites dans le document. Mais aucun outil n'est disponible pour retrouver les marques et les modifier (excepté le classique couper/coller). Seuls les numéros des pages figurant dans la table d'index permettent de retrouver les marques d'origine. Et, lorsque l'on travaille sur un écran, la manipulation des pages d'un document est généralement mal aisée.

1.3 Index électronique

L'approche des documents structurés [André 89] [Furuta 88] [Quint 90] et celle des hypertextes peuvent être combinées pour créer des documents dynamiques [Chen 88], [Bruza 90] [Furuta 90] [Stotts 91]. Cette convergence permet de trans-

former des tables d'index en véritable index électronique dans lequel les facilités de navigation, qui n'existent pas dans les documents traditionnels, sont développées. Pour l'auteur d'un document, le suivi des liens doit faciliter la saisie des informations utiles à l'indexation. Pour le lecteur d'un document électronique, le suivi des liens doit permettre d'accéder directement au contenu du document, évitant ainsi le défilement (*scroll* vertical) des pages du document.

C'est l'approche que nous avons choisie dans Grif. L'expérience d'intégration que nous décrivons au chapitre 2 permet d'obtenir des documents structurés dont les index sont manipulables comme des hypertextes. Les index de documents créés par Grif permettent à la fois la manipulation du document électronique à l'aide de son index et la production d'un index classique, comme dans les documents électroniques.

La figure 4 montre l'exemple d'un ouvrage volumineux qui est représenté par Grif sous la forme de plusieurs documents (des *Chapitres*) manipulables séparément. L'index général permet d'accéder au contenu de ces différents documents. Par ailleurs, chaque chapitre peut posséder son propre index. Le double clic sur une entrée de l'index général fait afficher le passage du document correspondant.

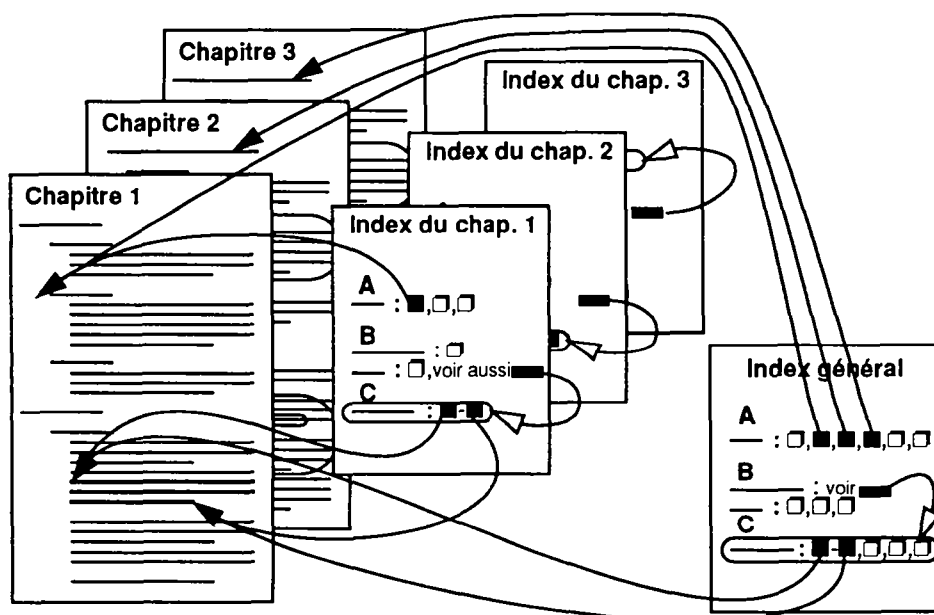


Figure 4: Index multi-documents

2 Les index électroniques dans Grif

Grif est un système de production de documents principalement destiné à la documentation scientifique et technique et, d'une façon plus générale, aux domaines où l'on manipule des documents fortement structurés. C'est un éditeur de documents structurés qui travaille à partir de structures génériques et de schémas de présentations [Quint 86]. Grâce aux structures génériques, les documents produits sont toujours «bien» structurés. Grâce aux schémas de présentation, les documents sont «bien» présentés. Mais Grif peut aussi être vu comme un éditeur d'hypertexte puisqu'il permet non seulement la gestion de liens entre éléments d'un document structuré, mais aussi la manipulation de liens entre documents (les hyper-documents).

2.1 Principe général de l'intégration dans Grif

2.1.1 Les documents structurés

La principale caractéristique de Grif est de manipuler des *documents structurés*. Pour Grif, un document est représenté comme une structure logique qui comprend des éléments tels qu'un *titre*, un *résumé*, des *chapitres*, *sections*, *paragraphes*, etc. Un méta-modèle permet de décrire les modèles de documents qui correspondent à chaque situation: un modèle de *Lettre*, un modèle d'*Article* et un modèle de *Livre* ne contiendront pas les mêmes éléments. Plusieurs articles (documents de la classe *Article*) ont un nombre différent de noms d'auteurs, de sections ou de paragraphes, mais ils sont construits sur le même modèle, celui de la structure logique générique de leur classe de document. Une structure générique spécifie les types d'éléments qui composent les documents d'une classe et la façon dont ces éléments peuvent s'organiser pour former un document. Ces structures génériques sont spécifiées, dans le langage S, sous la forme de *schémas de structure*.

Cette structure logique permet de produire automatiquement l'aspect graphique des documents. À chaque type d'élément de la structure logique générique est associé un ensemble de règles de présentation qui définissent l'aspect graphique de ce type d'élément. Un ensemble de règles de présentation, écrit en langage P, définissant l'aspect graphique de tous les types d'éléments d'une classe de documents est appelé un *schéma de présentation*. En appliquant ces règles aux éléments de la structure logique spécifique, on peut construire l'image du document. Ainsi, l'utilisateur est débarrassé du travail de mise en forme. En même temps, on assure une bonne homogénéité de la présentation et, lorsque c'est nécessaire, on peut garantir que le document est présenté selon un modèle imposé.

Un troisième langage, le langage T, permet d'exprimer des règles de traduction regroupées dans des *schémas de traduction*. Un schéma de traduction spécifie les règles de traduction ainsi que les chaînes de caractères à engendrer pour chaque élément d'un document, en fonction de son type et de sa position dans la structure

logique du document. Comme pour les schémas de présentation, on peut définir plusieurs schémas de traduction pour une même classe de documents, ce qui permet de traduire les documents d'une classe dans plusieurs formalismes différents.

Ainsi, un document produit par Grif peut être traduit selon un standard SGML. De plus, Grif peut être utilisé pour la saisie, la correction et la consultation de documents SGML. En effet, le modèle de document qui est à la base de Grif est conceptuellement proche du modèle des normes ODA [ISO 1989] ou SGML [ISO 1986]. Dans ces deux normes on a des structures logiques et physiques, des structures génériques et des structures spécifiques. Avec Grif, l'éditeur gère simultanément et de manière automatique la présentation du document (WYSISYG) et la représentation interne (en SGML) du document.

2.1.2 Les schémas d'extension

Afin de permettre la création d'index dans n'importe quel type de document (*Article, Thèse, Dossier*, etc.), les index ont été définis comme des **extensions**. Un schéma d'extension *ExtIndex* décrit les structures nécessaires à la construction des index et un schéma de présentation décrit la présentation de ces nouveaux éléments.

Ainsi, sans modifier les schémas existants, il est possible d'intégrer dans la structure d'un document les éléments structurés qui composent les index. Donc, quelle que soit la structure générique initiale du document, le mécanisme des extensions de schéma permet d'ajouter des index aux documents structurés manipulés par Grif. Plus généralement, ce mécanisme d'intégration de schéma permet d'ajouter les extensions utiles pour certaines applications: construire des index ou corriger un document, par exemple. Et, à la demande, ces extensions sont chargées dans le document.

2.1.3 Les liens hypertextes

Les structures logiques des documents sont essentiellement des structures arborescentes. Cependant, des liens non-hiérarchiques peuvent exister entre éléments d'arborescences. Ils sont représentés, dans le méta-modèle, par des *références*; une référence peut représenter soit un renvoi (à une note, à une figure, à une section, etc.), soit une inclusion d'un élément d'un document. Toutes les références, les renvois comme les inclusions, peuvent désigner des éléments logiques qui appartiennent au même document que la référence (*référence interne*) ou à un document différent (*référence externe*), donnant ainsi à Grif un aspect hypertexte [Quint 92a] [Quint 92b].

2.1.4 Les index électroniques

Comme tout élément logique, les index électroniques sont décrits dans les *schémas de structure* des documents (cf. 2.2) et leur aspect graphique est spécifié dans les *schémas de présentation* (cf. 2.3). Les *références* sont utilisées pour

décrire les liens entre les différents composants des index électroniques: les marques, leur descripteur et les entrées des tables d'index.

Certains liens sont créés explicitement par l'utilisateur, d'autres sont posés automatiquement par Grif. Le suivi des liens et la présentation des index sont ainsi contrôlés par l'éditeur Grif. L'utilisateur peut utiliser ces liens pour se déplacer dans le document à partir de l'index alphabétique qui est affiché sur l'écran. Cet index se présente comme un index traditionnel : c'est une liste alphabétique de termes et de références.

La construction des tables d'index (action **Construire les index**) est effectuée à la demande de l'utilisateur. Lorsque la pagination change, la reconstruction des tables est nécessaire pour que les numéros des pages soient mis à jour. Cependant, les liens entre table d'index et marques placées dans le document restent toujours valables, même si la pagination a changé.

Grif peut aussi gérer des super-documents: un super-document est une arborescence de documents. Cela permet de représenter des documents volumineux, tels que les documentations techniques ou les encyclopédies. Comme le montre la figure 4, chaque document contenu dans un super-document peut avoir une ou plusieurs tables d'index, le super-document peut aussi avoir plusieurs tables d'index: ces super-index électroniques permettent de faire des recherches sur l'ensemble des documents qui constituent le super-document.

2.1.5 Les modifications apportées à Grif

La version actuelle de l'éditeur Grif n'est pas encore suffisamment modulaire pour que l'intégration d'un nouvel outil puisse être réalisée sans modification de l'éditeur. Cependant, compte-tenu de l'expérience acquise lors de l'intégration d'un correcteur orthographique dans Grif [Frison 91], un certain nombre de facilités ont été introduites récemment pour permettre cette intégration: l'extension de schémas décrite en 2.1.2, la définition du bouton *Action* font partie de ces nouveautés introduites dans Grif pour faciliter l'intégration. En résumé, les modifications qui ont dû être apportées à Grif pour permettre la création de marques d'index et la génération des tables d'index sont les suivantes:

- modification du modèle de document: un nouveau type d'élément *Pair* a été introduit dans le modèle pour représenter des éléments couplés. Il est utilisé pour décrire les marques d'index dans les documents structurés.
- ajout des procédures de traitement de ce nouveau type *Pair* (insertion, suppression, sélection simultanée),
- modification de l'interface Grif: ajout de deux nouvelles commandes (*Actions*); l'une est utilisée pour demander la génération des tables d'index; l'autre permet de poser automatiquement des marques sur certains éléments et fonctionne comme une commande de *recherche*: cherche le type, l'attribut ou le contenu textuel de l'élément sur lequel poser des marques d'indexation,

- développement du mécanisme d'intégration de schémas pour autoriser des extensions de schémas, sans modifier les schémas existants.

La principale réalisation de ce projet a consisté à définir et mettre au point les nouvelles procédures chargées d'engendrer les tables d'index: la collecte des descripteurs, le tri, le regroupement des entrées, le calcul des références aux pages, des renvois et la construction du document structuré *index*.

Enfin, comme pour manipuler de nouveaux types de documents, la structure et la présentation des index ont été décrits dans des schémas. Grif propose pour cela deux langages: le langage S et le langage P. La structure des différents composants de l'index électronique (les marques, les descripteurs, les renvois, les tables d'index et leurs options de construction) fait l'objet d'un nouveau schéma d'extension *ExtIndex* écrit en langage S. Plusieurs schémas de présentation écrits en langage P sont proposés pour décrire la présentation de ces index.

2.2 Structure des index électroniques

Cette section décrit les nouvelles structures définies dans le schéma d'extension *ExtIndex*. Ces structures sont ajoutées aux structures des documents lors de la construction des index.

2.2.1 Paires de marques.

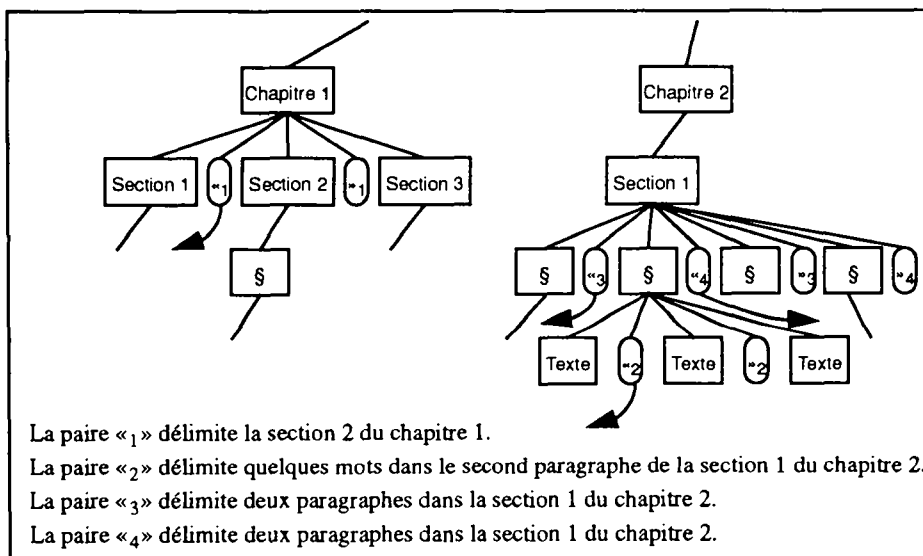


Figure 5: *Emplacement des marques dans la structure logique du document*

L'index électronique doit permettre à l'utilisateur de retrouver le passage qui l'intéresse dans le document. C'est pourquoi, à la différence des éditeurs classiques pour lesquels une marque d'indexation désigne un seul point dans le document, les marques d'indexation de Grif sont en fait des *paires de marques*.

Une paire de marques, placée autour d'un passage à indexer, désigne la zone du document qui sera sélectionnée à partir de la référence correspondante dans la table d'index.

La position de ces marques respecte toujours la structure arborescente du document : une paire de marque désigne une zone contiguë du document, c'est-à-dire une zone sélectionnable, par exemple un ou plusieurs paragraphes contigus, une section entière, deux sections consécutives, un chapitre entier, un titre ou un mot, etc. Mais, comme le montre la figure 5, ces paires de marques peuvent être imbriquées ou chevaucher : dans cet exemple, le 3^e paragraphe de la section 1 est à la fois « marqué » par la marque 3 et la marque 4 ; une partie du texte du second paragraphe porte aussi la marque 2.

Une paire de marques est donc considérée comme un élément de la structure logique du document. Elle est définie comme un élément de type *paire*. Ce type d'élément est constitué de deux éléments qui peuvent être désignés séparément dans les schémas en utilisant le mot-clé *First* ou *Second*, selon qu'il s'agit de désigner le premier ou le dernier élément d'une paire. Une paire est composée de deux éléments de même type qui seront toujours créés, déplacés ou détruits ensemble. Le méta-modèle a été étendu pour supporter ce nouveau type d'élément.

```
STRUCTURE EXTENSION ExtIndex;
DEFPRES ExtIndexP;
STRUCT
  Marque_Index
  (ATTR
    A_Vers_descripteur = REFERENCE(Descripteur);
    NuméroMarque = Integer) = PAIR;
....
EXTENS
  ROOT + (Marque_Index);
  Annexe - (Marque_Index);
ASSOC
  Descripteur (ATTR Vers_marque1 = REFERENCE(Marque_Index)) =
    BEGIN
      ....
    END;
END;
```

Figure 6: Structure des marques d'index
(extrait du schéma de structure des extensions pour index)

La figure 6 montre le contenu du schéma d'extension qui décrit ces marques : chaque paire est identifiée par un numéro, ce numéro est décrit dans le schéma de structure comme un attribut de type *entier* associé à la paire (*NuméroMarque*). À toute paire de marques est associé un descripteur dont le contenu précise un certain nombre d'informations qui seront utiles pour engendrer la table d'index

(cf. 2.2.2). Le lien entre la paire de marques et le descripteur est décrit comme un attribut de type *référence*: l'origine du lien est la paire de marques, l'extrémité du lien est le descripteur. Un second lien allant du descripteur vers la paire de marques est décrit comme un élément *référence*. Un double clic sur cet élément dans le descripteur permettra de sélectionner la paire de marque associée.

2.2.2 Descripteurs de marques d'index

Le contenu des descripteurs détermine ce qui apparaîtra dans l'index. La structure logique de ces descripteurs est imposée par le schéma de structure *ExtIndex* dont la figure 7 présente un extrait (écrit en langage S¹). Dans chaque descripteur, il est possible d'indiquer des niveaux multiples, de préciser l'importance de la référence, une catégorie et une glose éventuelle. L'utilisation des niveaux multiples permet de développer des index à plusieurs niveaux. Le texte de chaque entrée et sous-entrée est indiqué dans l'élément *Clé*. Selon la *valeur* indiquée, une présentation spéciale sera appliquée à la référence, par exemple le numéro de la page pourra être souligné ou en caractères gras (voir des exemples de présentation en 2.3). Par défaut, toutes les références sont créées avec une *valeur* normale.

```
STRUCTURE EXTENSION ExtIndex;
DEFPRES ExtIndexP;
STRUCT
....
EXTENS
....
ASSOC
  Descripteur (ATTR Vers_marque1 = REFERENCE (Marque_Index)) =
    BEGIN
      Clés (ATTR Valeur = faible, moyenne, forte)
        = LIST [1..3] OF (Clé = Contenu) WITH Valeur ? = moyenne;
      ? Sujet = TEXT;
      ? Sémantique = TEXT;
      ? Glose = Contenu;
    END;
END
```

Figure 7: Structure des descripteurs de marques
(extrait du schéma de structure des extensions pour index)

La *sémantique* permet de distinguer des homographes, c'est-à-dire des entrées d'orthographe semblable, mais dont la signification est différente. Par exemple, afin de distinguer la ville d'Orange, du fruit du même nom, on pourra préciser dans le descripteur associé à la ville que la *sémantique* est un *lieu* et dans le descripteur associé au fruit, on précisera que la *sémantique* est celle d'un *objet*. Ainsi, lors de la génération de l'index, les entrées de *sémantique* différente ne seront pas fusionnées.

1. En langage S, les éléments facultatifs sont précédés par un « ? ».

Il est possible de construire plusieurs index dans un même document. Le *sujet* permet de préciser la catégorie d'une entrée. Il sera ensuite possible de choisir des sujets différents pour composer différentes tables d'index en demandant à Grif de ne retenir que les entrées qui concernent ce (ou ces) sujet(s) (cf. description des options en 2.2.4). À la différence d'un éditeur comme FrameMaker qui prédéfinit les sujets (auteur, commentaire, etc.), Grif n'en propose *a priori* aucun. L'auteur est libre de définir les sujets qui seront les mieux adaptés au document qu'il compose.

Le commentaire placé à la suite du terme de l'entrée, avant les références aux pages, contient la *glose*¹ définie dans le descripteur.

Enfin, chaque descripteur est lié à une paire de marques d'index. Ce lien est décrit, dans le schéma de structure, par un attribut de type *référence*. Ainsi est-il possible, à partir d'un descripteur affiché dans une vue séparée du reste du document de retrouver directement la paire de marque correspondante dans le document. Les descripteurs de marque d'index sont affichés dans une vue séparée du document et éditables.

2.2.3 Descripteurs de renvoi

Dans une table d'index, les renvois sont des références qui n'indiquent pas un numéro de page, mais le texte d'une autre entrée (simple ou multi-niveau). En considérant ces renvois comme des éléments structurés d'un document, Grif peut contrôler automatiquement la cohérence des références croisées et leur appliquer un format de présentation spécifique. Cette approche améliore la qualité des tables d'index produites, à la fois dans leur contenu et dans leur présentation. Dans l'index engendré, ces renvois sont des liens entre des entrées d'index. Le suivi de ces liens facilite la consultation du document à partir de l'index (voir la figure 17).

Les renvois définis dans un document ne sont pas localisés de façon précise dans le document. Aucune marque ne leur est associée. Le méta-modèle de Grif permet de décrire de tels éléments comme des éléments associés, c'est-à-dire comme les notes ou les figures associées à certains documents. Pour ajouter des renvois dans un document, il suffit donc de créer un nouvel élément associé *Descripteur de renvoi* dans le document. De la même manière que les descripteurs de marques, les descripteurs de renvoi sont affichés dans une vue séparée du document et éditables.

Le contenu des descripteurs de renvoi détermine ce qui apparaîtra dans l'index. Les caractéristiques de ces descripteurs sont les mêmes que celles des descripteurs associés aux marques d'index : le texte de l'entrée et des sous-entrées éventuelles, la valeur de la référence, le sujet, la sémantique, la glose. Seule carac-

1. Une amélioration de la version actuelle consisterait à associer cette *glose*, non plus à un descripteur particulier, mais à un terme précis : ceci permettrait de conserver la *glose*, même si le descripteur est supprimé et éviterait l'ambiguïté des commentaires multiples toujours possibles, si plusieurs descripteurs de même clé contiennent une *glose*.

```

STRUCTURE EXTENSION ExtIndex;
DEFPRES ExtIndexP;
STRUCT
....
EXTENS
....
ASSOC
  Descripteur_de_renvoi =
    BEGIN
      CClés (ATTR Valeur = faible, moyenne, forte)
      = LIST [1..3] OF (CClé = Contenu) WITH Valeur ? = moyenne;
      ? Sujet;
      ? Sémantique;
      ? Glose;
      Renvoi = REFERENCE(Clés);
    END;
END

```

Figure 8 : *Structure des descripteurs de renvoi*
(extrait du schéma de structure des extensions pour index)

téristique supplémentaire: le texte de l'entrée (et des sous-entrées éventuelles) référencées par ce renvoi. Dans la structure logique, cette caractéristique n'est pas considérée comme un texte, mais comme une *référence*. Cette approche évite les erreurs de saisie et permet une vérification immédiate de l'existence de l'entrée référencée: on affiche [?] dans le cas de renvoi vide.

2.2.4 Options de construction des tables

Grif ne peut construire d'index que si les caractéristiques en ont été préalablement définies. Les caractéristiques des tables à construire sont décrites dans des éléments associés au document. Elles sont affichées dans une vue séparée du document et éditables avec Grif.

Comme tout élément du document, ces caractéristiques ont une structure logique décrite dans le schéma de structure (extension pour index). S'il y a plusieurs tables d'index définies pour un même document, chacune d'elles fait l'objet d'une description spécifique. Il y a donc autant d'options associées au document que de tables d'index à construire pour ce document. Ainsi, bien que la structure générique des tables d'index soit la même, quel que soit le contenu de la table, certaines variantes peuvent être définies dans ces options pour influencer sur la structure, la présentation et le contenu de ces tables:

- le titre de la table d'index;
- un paragraphe introductif (de contenu libre) permet de décrire, par exemple, le contenu ou la présentation de la table d'index, pour le futur lecteur;

- l'ordre de tri des entrées (avec ou sans séparation des majuscules et minuscules, la position des nombres, etc.);
 - la présentation des renvois: vers les numéros de page ou vers les numéros de section;
 - la typographie des entrées: avec ou sans capitale initiale ou respectant la typographie d'origine;
 - la présence de titres de groupe: A, B....Z;
 - la fusion de certains groupes d'entrées: J-K ou X-Y-Z, par exemple;
 - le sujet des descripteurs sélectionnés pour figurer dans cette table d'index;
- la présentation spécifique des entrées traitant d'un sujet donné.

```

STRUCTURE EXTENSION ExtIndex;
DEFPRES ExtIndexP;
STRUCT
....
EXTENS
....
ASSOC
....
Option = BEGIN
  Idx_Titre (ATTR Tri = alpha, num, typo, personnel;
             Pres_Renvoi = Par_page, Par_section;
             Pres_Terme = MAJUSCULES, Minuscules, Une_capitale,
             Des_capitales, Tel_quel;
             Pres_Groupe = Visible, NonVisible) = Titre_Index
    WITH      Tri ? = alpha,
             Pres_Renvoi ? = Par_page,
             Pres_Terme ? = Une_capitale,
             Pres_Groupe ? = Visible;
  ? Idx_Sujets = LIST [0..4] OF (Idx_Sujet
    (ATTR Pres_Sujet = romain, gras, italique, italgras) = TEXT
    WITH      Pres_Sujet ? = romain);
  ? Idx_Intro = Intro_Index;
  ? Idx_Groupes = LIST OF (Idx_Groupe =
    BEGIN
      Idx_CarGroupe = LIST [2..*] OF (Idx_CarElem = TEXT);
    END);
  END
  - (Marque_Index);
END

```

Figure 9: Structure des options de construction
(extrait du schéma des extensions pour index)

Ces caractéristiques sont décrites dans le schéma de structure *ExtIndex* (figure 9), soit comme des *attributs*, soit comme des éléments textuels. L'utili-

sation d'attributs de type *énumération* permet d'éviter des erreurs de saisie, en limitant les choix des utilisateurs à une liste de valeurs permises.

Dans le cas des super-documents, les options de construction des tables d'index sont définies dans le super-document, alors que les marques et les descripteurs des marques restent dans chacun des documents qui composent le super-document.

2.2.5 Table d'index

Les tables d'index d'un document peuvent avoir un nombre quelconque d'entrées et de références. Mais toutes ces tables sont construites selon le même modèle décrit dans le schéma de structure (extensions pour index). La structure d'une table d'index est celle d'une liste d'entrées précédée par un titre et éventuellement un paragraphe d'introduction. Les entrées peuvent être groupées et précédées d'un titre de groupe. Chaque entrée est constituée d'un terme, d'une glose éventuelle et d'une liste de références. Certaines entrées sont hiérarchisées sur plusieurs niveaux; dans ce cas, un terme d'indexation est défini pour chaque niveau. Les listes de références sont constituées de références internes au document (généralement, le numéro de page sert de référence), suivies éventuellement de références internes à l'index (les renvois vers une autre entrée de l'index).

La description, en langage S, de la structure logique générique des tables d'index construites par Grif est la suivante:

```

STRUCTURE EXTENSION ExtIndex;
DEFPRES ExtIndexP;
STRUCT
  Car_Index = BEGIN
    Titre_Car = LIST [0..*] OF (Car = TEXT);
    Contenu_Car = LIST [0..*] OF (Elem_Index);
  END;
  Elem_Index = BEGIN
    Entrée;
    ? Suite_Entrée = LIST [0..*] OF (Entrée1);
  END;
  Entrée = BEGIN
    Terme (ATTR Niveau = roamin, gras, italique, italgras) = TEXT;
    ? Commentaire = Glose;
    ? Liste_des_ref = BEGIN
      ? Liste_directe = LIST [0..*] OF (RefDirecte);
      ? Liste_croisée = LIST [0..*] OF (RefCroisée);
    END;
  END;
  RefCroisée = BEGIN
    RefCr (ATTR Force=faille, normal, fort) = REFERENCE(Entrée);
    Vers_croisée = REFERENCE (CClés);
  END;
  RefDirecte = BEGIN
    ? NoteNum = TEXT;

```

Grif et les index électroniques

```

PageNum (ATTR Force) = CASE OF
  PageUnique (ATTR Vers_marque
              = REFERENCE (First Marque_Index)) = TEXT;
  PageIntervalle = BEGIN
  PageDebut (ATTR Vers_marque) = TEXT;
              Pagefin (ATTR Vers_marque_fin
                      = REFERENCE(Second Marque_Index)) = TEXT;
              END;
  END;
Vers_descripteur = REFERENCE(Clés);
END;
Entrée1 = BEGIN
  Entrée;
  ? Suite_entrée1 = LIST [0..*] OF (Entrée2);
  END;
Entrée2 = BEGIN
  Entrée;
  ? Suite_entrée2 = LIST [0..*] OF (Entrée3);
  END;
Entrée3 = Entrée;
....
EXTENS
....
ASSOC
  I_Table(ATTR Pres_Ref = Par_page, Par_section;
          Pres_Car = Visible, NonVisible) = BEGIN
  Titre_Index = Contenu;
  ? Intro_Index = LIST OF (Paragraphe);
  Contenu_Index = LIST [0..*] OF (Car_Index);
  END;
....
END

```

Figure 10: Structure de la table d'index
(extrait du schéma des extensions pour index)

Les éléments nouveaux par rapport à un index traditionnel sont les suivants:

- les références internes au document sont des liens vers les marques d'indexation placées dans le document: un double clic sur cette référence, dans la vue index, fait afficher, dans la vue intégrale du document, le passage marqué correspondant.
- les références internes à l'index sont des liens vers une autre entrée de l'index. Ainsi, par exemple, si l'entrée *Pays-Bas* est suivie du renvoi *voir Hollande*, un double clic sur *voir Hollande*, fait afficher l'entrée *Hollande*, suivie de la liste de références associées à cette entrée. Il suffira alors de choisir l'une de ces références pour faire afficher un passage concernant la *Hollande*.
- les liens avec les descripteurs (de marques et de renvois) sont placés à droite de chaque référence. Ainsi, afin de modifier le contenu d'une entrée

de l'index, il suffit de cliquer à droite de la référence afin de faire afficher immédiatement le contenu du descripteur. Ce descripteur peut alors être édité avec Grif.

Dans le cas d'un super-document, les tables d'index, ainsi que les options sont définies dans le super-document. Les liens entre index et marques et entre index et descripteur sont des *références externes*.

2.3 Présentation des index électroniques

Le modèle de document adopté par Grif permet de séparer clairement la présentation des documents de leur structure et de leur contenu. L'organisation logique d'un document est utilisée pour effectuer sa présentation. Comme la structure logique, la présentation des documents est définie à l'aide d'un langage (le langage P) dans un *schéma de présentation*. L'utilisation des schémas de présentation permet d'assurer une présentation homogène des documents. Il est possible d'écrire plusieurs schémas de présentation différents pour une même classe de documents. Cela permet d'offrir aux utilisateurs le choix de l'aspect graphique des documents qui convient le mieux à leur type de travail.

Pour les index électroniques, un aspect graphique peut mieux convenir à l'affichage et à la manipulation des index et un autre à l'impression des tables d'index et du document. Dans le document imprimé, par exemple, les marques d'index, les descripteurs, les renvois et les liens ne sont pas visibles. Les utilisateurs pourront donc choisir un schéma de présentation différent selon qu'il s'agit de consulter un document et de l'imprimer ou bien de construire les tables d'index de ce document (voir la figure 17).

La présentation de toutes les tables d'index d'un même document est conforme au schéma de présentation défini pour ces tables. De plus, certains attributs des descripteurs et des options conditionnent la présentation des tables d'index : des attributs sont posés sur certains éléments des tables d'index (lors de la construction de ces tables) pour en modifier la présentation. Par exemple, l'attribut *Force*, appliqué aux références aux pages, permet d'obtenir de l'italique ou du gras pour présenter les numéros des pages référencées dans l'index; l'attribut *Pres_Car*, appliqué à une table d'index, permet de rendre visible ou invisible les titres des groupes d'entrées.

2.4 Construction des tables d'index

L'action **Construire les index** a pour effet de créer des tables d'index dans le document courant. Ces tables ont une structure conforme à la structure générique décrite en 2.2.5. Ce sont des documents structurés, contenant des références vers les marques du document courant et vers les descripteurs de ces marques et des renvois.

Tous les descripteurs (de marque et de renvoi) traitant du sujet demandé dans les options sont sélectionnés pour être référencés dans les tables d'index. Seuls les

descripteurs «complets» (ayant *clé* et *renvoi*, dans le cas des descripteurs de renvoi) sont pris en compte.

Pour l'utilisateur, cette construction est immédiate: les nouvelles tables remplacent les anciennes dans la vue *Index*. Mais le fonctionnement interne peut être décomposé en trois étapes:

- (1) sélection des descripteurs concernés et remplissage d'une table intermédiaire contenant les clés de niveau 1, 2 et 3, la sémantique, le numéro de la page¹ de la première marque d'index, le numéro de la page de la seconde marque d'index, un pointeur vers le descripteur. Dans le cas de descripteurs de renvoi, les numéros de page sont nuls;
- (2) tri de cette table intermédiaire: dans l'ordre alphabétique ou alphanumérique des clés de niveau 1 à 3, en tenant compte des homographes, (définis dans le descripteurs par le champ *sémantique*); le résultat de ce tri est une liste ordonnée des indices dans la table intermédiaire;
- (3) création de la table d'index: une première passe permet de construire l'arbre abstrait, conforme au schéma de structure, contenant toutes les entrées décrites dans la table intermédiaire obtenue en (1). Une seconde passe est nécessaire pour remplir les références croisées, en particulier les références en avant dans la table d'index.

S'il y a plusieurs tables d'index à construire, les étapes (1), (2) et (3) sont reproduites pour construire chaque table.

1. ou le numéro de la section (si l'option de construction le stipule).

3 Interface utilisateur des index de Grif

Grif est un éditeur interactif qui permet de manipuler directement les documents sous leur forme graphique : la représentation graphique des documents, construite à partir des schémas de présentation, est affichée sur l'écran. L'utilisateur peut, au moyen de la souris, du clavier, au travers de fenêtres, de menus et de formulaires de dialogue éditer un document. Les effets des commandes sont répercutés immédiatement sur la représentation graphique du documents.

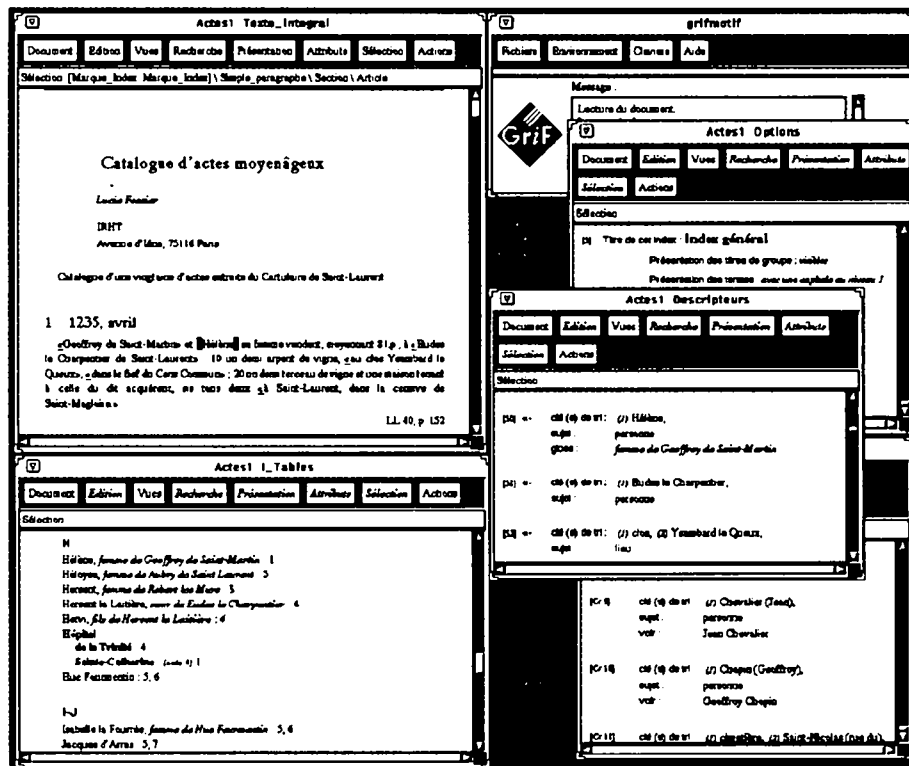


Figure 11 : Les vues Grif : Texte_intégral, Table d'index, Options, Descripteurs de marque, Descripteurs de renvoi

3.1 Chargement des extensions

La pose des marques d'index et la construction des tables d'index ne peut avoir lieu dans un document que si les extensions pour index ont été préalablement chargées dans ce document, quel que soit le schéma de structure de ce document (*Article*, *Chapitre*, *Dossier* ou autre). Le chargement de ces extensions est effectué lors du premier appel de l'action **Poser des marques** ou de l'action **Construire les index** dans un document. Ensuite, l'insertion des marques dans ce document est autorisée conformément aux règles décrites dans le schéma des extensions pour index.

Lors de l'ouverture d'un document contenant déjà des marques ou des options de construction d'index, le chargement des extensions nécessaires au bon fonctionnement des index est automatique. Le chargement¹ des extensions est donc implicite dans un document contenant des éléments décrits dans le schéma d'extension.

3.2 Vues

La représentation graphique du document manipulé par Grif est généralement composée de plusieurs vues. Chaque vue montre tout ou partie des éléments contenus dans le document. Le contenu de chaque vue est spécifié dans les *schémas de présentation*. Un modèle du document *Article*, par exemple, est généralement manipulé au travers de trois vues: la vue *Texte intégral*, la vue *Table des matières* et la vue des *Citations bibliographiques*.

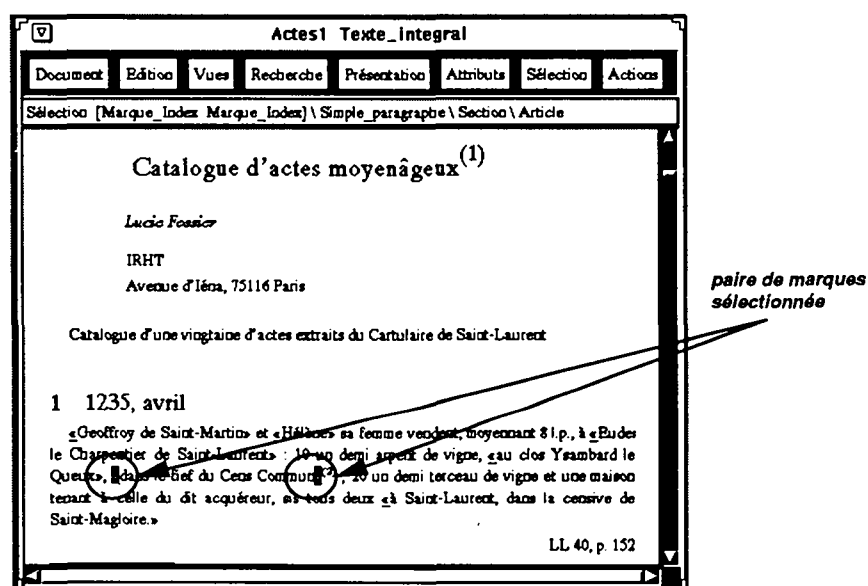


Figure 12: Affichage des marques d'index dans la vue *Texte_intégral*

Pour présenter les index, plusieurs vues nouvelles du document sont définies dans le *schéma de présentation* des extensions. Ces vues sont ajoutées aux autres vues du document : la vue des *Tables d'index*, celle des *Descripteurs de marque*, celle des *Descripteurs de renvoi* et celle des *Options de construction* des tables d'index.

La figure 11 montre les vues utilisées lors de la construction des tables d'index: en cliquant sur les marques sombres (affichées plus distinctement en couleur sur l'écran) de la vue *Texte_intégral*, le descripteur correspondant est

1. Aucune commande ne permet d'extraire ces éléments du document.

affiché et sélectionné dans la vue *Descripteur*. La vue des *Tables d'index* permet, comme la vue *Table des matières*, un accès direct au document par son contenu. Il suffit de cliquer sur une référence (numéro de page ou de section) ou sur un renvoi. Les vues des *Descripteurs* et des *Options* ne présentent d'intérêt que pour l'auteur du document lorsqu'il effectue la mise au point des tables d'index.

3.3 Marques

Les marques d'index peuvent être placées dans un document aux emplacements autorisés par le schéma de structure du document. Leur pose peut être interdite dans certaines parties du document si cela est spécifié explicitement dans le schéma de structure *ExtIndex*. Leur position est compatible avec la structure du document : une paire de marque encadre un élément ou une suite d'éléments frères dans l'arborescence. En fait, les marques peuvent être posées autour de n'importe quel élément *sélectionnable* d'un document. Une paire de marques est toujours sélectionnée comme un seul élément : il est impossible de ne sélectionner qu'une seule marque séparément (figure 12).

À tout moment, l'éditeur de document Grif maintient à jour le menu des éléments qu'il est possible d'insérer dans le document : les marques d'index apparaissent donc dans ce menu lorsque l'insertion est autorisée conformément au schéma de structure.

Figure 13: Formulaire d'insertion automatique de marques d'index

Deux façons de procéder sont proposées dans Grif pour poser ces marques. La première consiste à utiliser la commande **Insérer** pour placer les marques autour d'un passage préalablement sélectionné. La seconde consiste à lancer une commande de recherche spécialisée pour **Insérer des marques**. Cette commande

recherche dans le document courant des éléments selon leur contenu (texte), leur type ou l'attribut qu'ils portent, les sélectionne et pose des marques d'index autour d'eux, en demandant éventuellement confirmation à l'utilisateur. Cette méthode de pose permet de s'appuyer sur la structure logique du document: par exemple, pour mettre dans l'index les noms des *auteurs*, les *titres de section* ou les mots portant un attribut «*Importance = forte*», lorsque ces types sont définis dans le schéma de structure du document..

Quelle que soit la méthode utilisée, l'éditeur Grif est toujours guidé par la structure logique générique du document pour contrôler, à tout moment, la sélection et donc aussi le positionnement correct des marques. Toute création de marque construit automatiquement le descripteur associé.

3.4 Descripteurs de marque

Le descripteur de marque est créé automatiquement par Grif lors de la pose d'une paire de marques d'index. Un descripteur est toujours solidaire des marques qui l'ont créé: il ne peut être supprimé que par une commande de suppression des marques (sélectionnées). Il n'est pas possible non plus de créer directement un descripteur. Toutes les commandes d'édition appliquées aux marques s'appliquent en même temps au descripteur associé (copier, couper, coller).

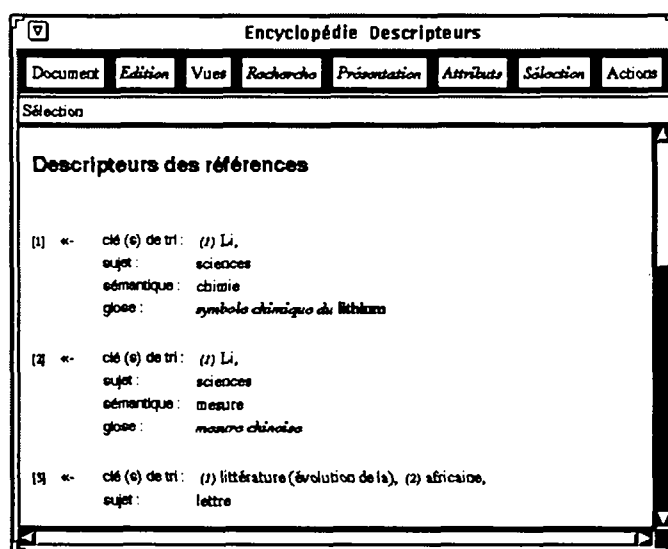


Figure 14: Saisie des descripteurs dans la vue *Descripteurs*

Le contenu d'un descripteur peut être édité avec Grif, comme les autres éléments du document: une valeur initiale est affectée automatiquement à la clé (c'est le premier mot de la zone marquée). Cette clé peut être modifiée. Des sous-clés, une glose, un sujet, une sémantique peuvent être ajoutée / supprimés / modifiés. Mais, seuls les descripteurs ayant au moins une clé définie (non vide)

seront retenus pour construire les tables d'index. Les autres champs sont facultatifs.

La **sémantique** permet de distinguer les homographes¹: ainsi, deux termes identiques, mais de signification différente, par exemple, *Orange* (fruit) et *Orange* (ville) ne feront pas l'objet d'une fusion dans la table d'index si leur descripteur précise une sémantique différente, à savoir, *fruit* pour l'un et *ville*, pour l'autre, par exemple. C'est donc à l'auteur du document de préciser dans les descripteurs la sémantique des termes qui ne doivent pas être fusionnés, même si leur orthographe est semblable.

Le **sujet** est utilisé pour filtrer les entrées et construire des index spécialisés. Il ne fait l'objet d'aucun contrôle. Il est donc conseillé de faire un choix pertinent des noms de sujet afin de faciliter ce filtrage et d'éviter les fautes de frappe. De plus, un descripteur ne peut avoir qu'un seul sujet. En conséquence, ce sujet doit être précis pour faciliter les filtrages ultérieurs éventuels. Par exemple, le sujet « lieu » appliqué à la fois à des noms de villes et des noms de pays ne permettra pas de construire des index distincts pour les villes et pour les pays. Par contre, le choix de deux sujets « ville » et « pays » n'empêchera pas de construire un index global contenant à la fois les villes et les pays: une table d'index peut contenir des termes traitant de plusieurs sujets (le nombre maximum de sujets est défini dans le schéma de structure des extensions pour index).

Lorsque plusieurs descripteurs de même clé ont une **glose**, seule l'une d'entre elle apparaîtra en commentaire dans la table d'index. Dans la table d'index, le commentaire est une *inclusion* de la glose d'un descripteur. Ainsi, toute mise à jour de cette glose apparaît immédiatement dans les commentaires des tables d'index correspondants: Grif contrôle à tout moment le contenu des inclusions.

Par contre, toute modification des autres caractéristiques d'un descripteur ou tout ajout ou suppression d'un descripteur (lors de la pose ou destruction de marques) sera sans effet immédiat sur le contenu des tables d'index. À tout instant, Grif vérifie cependant que les liens de type *référence* existent. En conséquence, si une paire de marques et son descripteur sont supprimés, le descripteur d'un renvoi (voir 3.5) vers ce descripteur supprimé apparaîtra avec un symbole indiquant une référence vide (par exemple [?]). Il faudra demander explicitement la reconstruction des tables pour voir les effets de ces modifications: voir l'action **Construire les index** décrite en 3.7.

3.5 Descripteurs de renvoi

Aucune interface spéciale n'est nécessaire pour manipuler les *Descripteurs de renvoi*: il suffit de demander l'ouverture de la vue *ad hoc*. Comme les autres éléments associés au document, ils peuvent être ajoutés, modifiés ou supprimés

1. Le nombre d'homographes différents est actuellement limité à trois (c'est-à-dire, trois valeurs de la *sémantique*, au maximum pour une même clé).

avec Grif en utilisant une simple commande d'insertion, d'édition ou de suppression dans la vue des Descripteurs de renvoi.

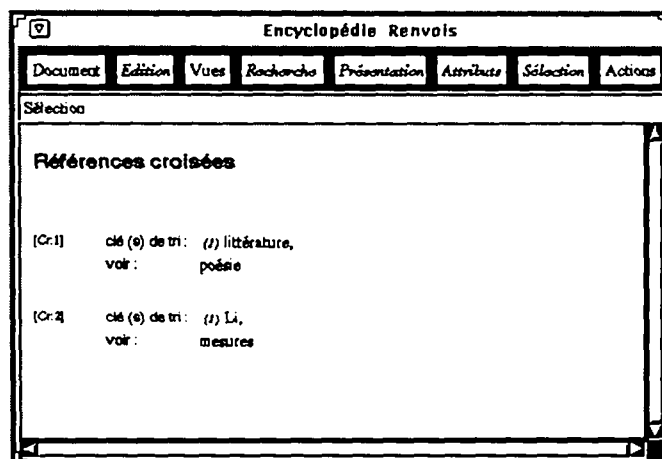


Figure 15: Saisie d'un renvoi dans la vue *Renvois*

Comme les descripteurs de marque, les descripteurs de renvoi peuvent avoir des sous-clés, un sujet, une sémantique et une glose. Aucune valeur initiale n'est définie par Grif pour la clé. Un descripteur de renvoi sera pris en compte lors de la construction des tables d'index si :

- (a) la clé est définie (non vide),
- (b) le renvoi est défini (renvoi vers un descripteur existant dans la table).

Par exemple, un renvoi vers un descripteur existant dans le document, mais n'ayant pas un sujet choisi pour la table à construire, ne sera pas retenu: seules les renvois internes à une table sont créés dans une table d'index.

On remarquera que, lorsqu'un descripteur de marque (pointé par un renvoi) est supprimé, le descripteur du renvoi vers ce descripteur supprimé apparaît avec un symbole indiquant une référence vide (un [?] est affiché sur l'écran à la place de la référence).

3.6 Options de construction

Aucune interface spéciale n'est nécessaire pour manipuler les *Options de construction* : il suffit de demander l'ouverture de la vue *ad hoc*. Comme les autres éléments associés au document, ils peuvent être ajoutés, modifiés ou supprimés avec Grif en utilisant une simple commande d'insertion, d'édition ou de suppression dans la vue des *Options de construction*.

Une liste d'option doit contenir obligatoirement un titre (**Idx_Titre**) pour être prise en compte lors de la construction des index (voir figure 6). Les autres options prennent des valeurs définies par défaut dans le schéma de présentation des extensions pour index : le tri des entrées sera alphanumérique avec fusion des majuscules et minuscules (attribut **Tri** = alpha), les termes seront présentés avec

une capitale initiale (attribut **Pres_Terme** = Une_Capitale), les références renverront au numéro de page (attribut **Pres_Renvoi** = Par_page), les titres des groupes seront visibles (attribut **Pres_Groupe** = visible) et tous les sujets seront retenus (**Idx_Sujet** vide) et présentés en romain (attribut **Pres_Sujet** = romain). Le regroupement de lettres peut être demandé en ajoutant un élément (facultatif) **Idx_CarGroupe**. L'introduction (**Idx_Intro**) est facultative.

Le titre de la table et l'introduction sont copiés par *inclusion* dans la table d'index : ces sont des copies vivantes de texte. En conséquence, toute modification de ces éléments dans les options a un effet immédiat sur la table d'index. Par contre, toute modification des autres options, tout ajout ou suppression d'une nouvelle liste d'options sera sans effet immédiat sur le contenu des tables d'index. Pour en voir l'effet, il suffit de lancer à nouveau l'action **Construire les index**.

3.7 Construction des tables d'index

La construction des tables d'index est une commande nouvelle offerte par Grif. Cette commande (Action **Construire les index**) ne peut être appliquée qu'aux documents préalablement ouverts et sélectionnés contenant déjà des marques d'index¹. Si la construction des tables concerne un super-document, des liens entre les documents contenus dans le super-document et la table d'index du super-document seront créés : pour l'utilisateur, il n'y a aucune différence entre les index de documents et les index de super-documents.

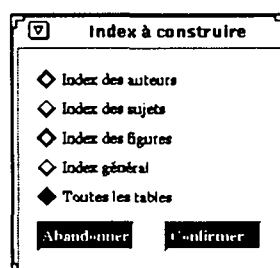


Figure 16: Formulaire de choix des tables d'index à construire

Cette commande permet de construire ou de reconstruire l'ensemble des tables d'index d'un document ou seulement certaines d'entre elles. Les titres des tables d'index définies dans la vue *Options* sont repris dans le formulaire de construction (voir figure 16). Il est nécessaire de demander la reconstruction des tables d'index si les descripteurs ou les options ont été modifiés. En effet, il n'y a pas, dans la version actuelle de Grif, de mise à jour automatique des tables d'index. Une telle

1. Le chargement des extensions est implicite lors du premier appel d'une action **Construire les index** ou **Poser des marques** ou lorsque le document contient déjà des marques d'index, des descripteurs de renvoi ou des options. Plus généralement, si un document contient des éléments définis dans un schéma d'extension, ce schéma d'extension est automatiquement chargé lors de l'ouverture du document.

mise à jour, lorsque les tables d'index sont volumineuses, demanderait un calcul complexe qui n'est pas justifié dans un environnement interactif. Par contre, tous les liens entre les marques et les descripteurs sont maintenus à jour (par le mécanisme des *références*). Les éléments copiés par *inclusion* sont également mis à jour par Grif, comme cela a été indiqué, dans les sections précédentes. Cela concerne les éléments suivants: le *titre* et l'*introduction* de l'index, le *commentaire* des termes.

Enfin, l'édition interactive de la table d'index est interdite par le schéma de présentation (table d'index protégée en *read_only*). En l'absence d'une telle protection, il serait possible de modifier le contenu de la table d'index; toutefois, cette solution présente l'inconvénient suivant: en cas de reconstruction des tables d'index, toutes les modifications de la table d'index sont perdues. Il est donc préférable de modifier, si nécessaire, les descripteurs ou les options afin de pouvoir reproduire les changements dans les constructions ultérieures des tables d'index. Et cela peut se faire simplement en suivant les liens électroniques décrits en 3.8.

En cas de reconstruction, les anciennes tables et leurs liens avec les éléments du document sont supprimés. Les nouvelles tables sont créées, dans la vue des *Tables d'index*, ainsi que les nouveaux liens entre ces tables d'index et les marques et entre ces tables et les descripteurs. Ainsi, alors que la structure logique du document est stable, la structure des index créés par Grif dans un document structuré est dynamique.

3.8 Liens électroniques

Dans un document Grif, les liens ne sont pas visibles. Afin de pouvoir identifier les origines des liens existants dans un document, il faut donner à ces éléments un aspect graphique caractéristique différent des autres éléments du document.

<p>B Baudoin Boucel, <i>bourgeois de Paris</i> : (note 11) s. 19 Baudoin de Cligny : s. 9 Bégan (Guillaume) : voir Bégan (Guillaume) Bégan (Jean) : voir Jean Bégan Bourgeois, <i>bru de Guillaume Galopin</i> : s. 4</p>	<p>B Baudoin Boucel, <i>bourgeois de Paris</i> : (note 11) s. 19 -> Baudoin de Cligny : s. 9 -> Bégan (Guillaume) : voir Guillaume Bégan -> Bégan (Jean) : voir Jean Bégan -> Bourgeois, <i>bru de Guillaume Galopin</i> : s. 4 -></p>
renvois internes et références aux sections	descripteurs de marques (->) et de renvoi (-x->)

Figure 17: Affichage d'une table d'index (extrait)

Cette présentation est définie dans le *schéma de présentation*, soit par une couleur différente, soit par une typographie différente. Ensuite, un double clic est la façon la plus simple de suivre ces liens: un double clic sur l'origine d'une *référence* (*attribut* ou *élément*) est considéré par l'éditeur comme une demande d'affichage de l'extrémité du lien concerné.

La figure 17 montre deux présentations de la table d'index: dans la vue de gauche, seules apparaissent les références croisées et les références aux sections;

c'est la vue la plus proche de la version imprimée de la table d'index; dans la vue de droite, les liens vers les descripteurs sont visibles et représentés par une flèche (en couleur sur l'écran); en suivant ces liens (double clic sur une flèche), l'utilisateur peut accéder directement au descripteur de marque ou au renvoi concerné).

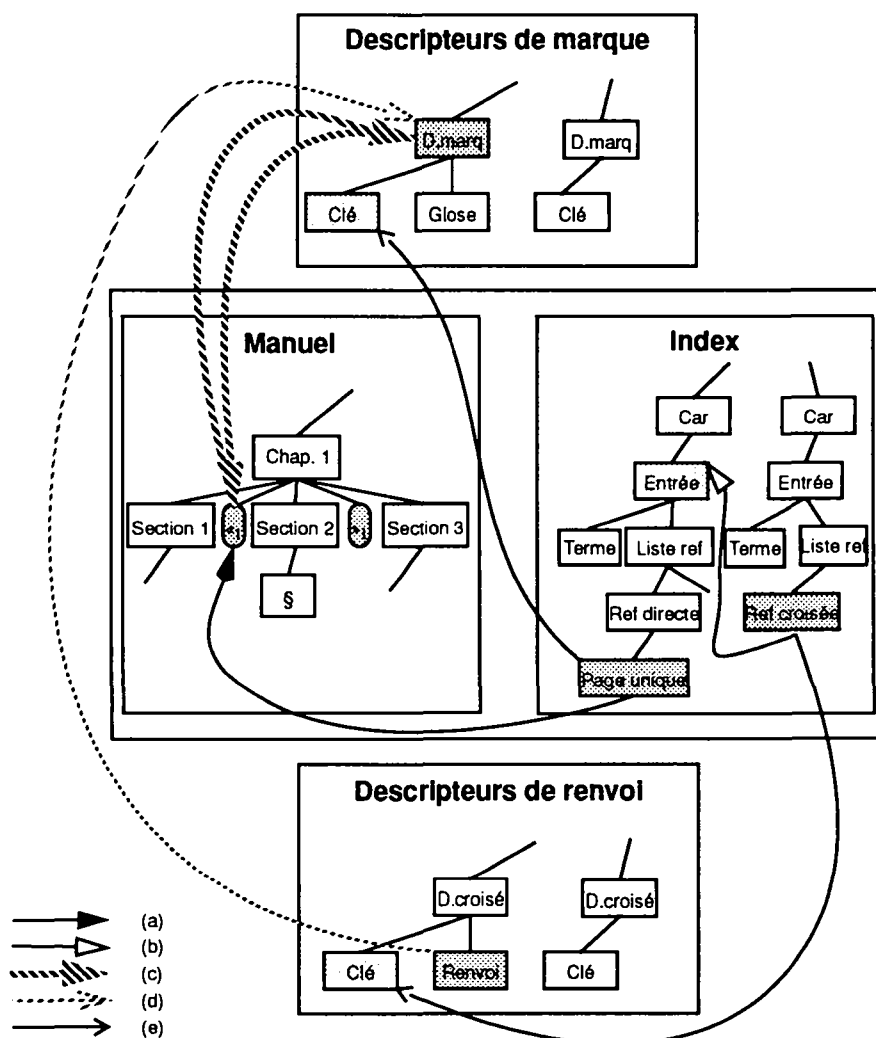


Figure 18: Utilisation des éléments de type *référence* dans les index

Les index électroniques coordonnent plusieurs types de liens (cf. figure 18:

- (a) les liens entre les tables d'index et les passages marqués du document,
- (b) les renvois internes dans une table d'index,
- (c) les doubles liens entre les marques et leur descripteur,
- (d) les liens définissant les renvois: entre un descripteur de renvoi et un descripteur de marque,

(e) les liens entre la table d'index et les descripteurs.

S'il s'agit uniquement de consulter un document électronique (approche *lecteur*), seuls les liens (a) et (b) sont utilisés. Par contre, pour mettre au point les index, au cours de l'édition d'un document (approche *auteur*), les liens (c), (d) et (e) sont aussi très utiles.

3.8.1 Consultation d'un document électronique

L'approche du **lecteur** consiste à rechercher dans le document électronique les passages intéressants. Avec un éditeur de documents structurés tel que Grif, cette recherche peut s'effectuer de deux manières: soit à partir de la table des matières, soit à partir des tables d'index. Le mécanisme de synchronisation des vues permet, par un simple clic sur un élément dans la vue *Table des matières* de voir s'afficher cet élément dans la vue *Texte intégral* du document.

Dans le cas des index, en utilisant la vue des *Tables d'index*, les possibilités sont plus étendues. Si le lecteur trouve directement, dans l'une des listes alphabétiques des index, le terme qui l'intéresse suivi d'une ou plusieurs références à des pages du document, alors il lui suffit de cliquer successivement sur chacune des références pour voir s'afficher les passages concernés. Si le terme qu'il cherche est suivi d'un ou plusieurs renvois, alors un double clic sur les termes du renvoi lui permet de visualiser une nouvelle liste des références, etc. Si le terme qu'il cherche ne figure dans aucun index, il doit essayer d'en trouver des synonymes et consulter à nouveau l'index pour essayer de trouver un terme approchant.

Dans un document volumineux, la recherche dans une table d'index est plus efficace que la recherche dans l'ensemble du document. Le mode d'accès au document «par le contenu» s'ajoute ici au mode d'accès «par la structure» offert par la table des matières dans un document structuré et à la «lecture séquentielle» du document.

3.8.2 Création d'un document électronique

L'approche de l'**auteur** est sensiblement différente. Bien sûr, l'auteur peut également, en cours d'édition, construire des tables d'index et les utiliser pour accéder au document. Mais il peut aussi utiliser les liens avec les descripteurs pour mettre au point le contenu de ces descripteurs et améliorer les tables d'index.

Pour une manipulation plus aisée, les liens entre marques et descripteurs sont doubles ; ils peuvent être suivis dans n'importe quel sens: il suffit d'un double clic sur une marque pour visualiser et sélectionner le descripteur associé; inversement, à partir du descripteur, il suffit d'un double clic sur le symbole (->) représentant le lien, pour visualiser la première marque associée et le début du passage correspondant dans la vue *Texte intégral* du document. Ces liens permettent de passer rapidement de la vue des *Descripteurs de marques* aux marques elles-mêmes, sans utiliser la commande de recherche. Ils sont très utiles pour la mise au point des descripteurs.

4 Résultats et perspectives

4.1 Premières expérimentations

Les premières expérimentations de ces index électroniques sont faites à l'Irisa sur quelques documents très différents, tant par leur structure que par leur contenu :

- un chapitre d'un livre sur l'industrie du lait;
- un catalogue d'actes moyen âgeux (figures 11, 12 et 17);
- le rapport d'activité du laboratoire (en cours).

Ces expériences permettent de constater que l'utilisation de Grif facilite le travail lorsqu'il s'agit de créer des tables d'index importantes. Cependant, le choix des clés reste primordial pour assurer la qualité d'une table d'index. Ce choix et celui de l'emplacement exact des marques dans le document reste difficile: c'est ce choix qui détermine les termes à partir desquels la lecture d'un document pourra être orientée, à l'aide d'une table d'index.

Avec Grif, une paire de marques peut être placée autour de n'importe quel élément du document. Cependant, la pose des marques est souvent effectuée **autour d'un seul mot**, préalablement sélectionné. Cette façon de procéder est la plus simple et correspond à la manière traditionnelle d'indexer des documents, en soulignant à la main les mots importants d'un texte. L'intérêt de placer ces marques autour des zones plus étendues n'apparaît que lors de la consultation de la table d'index électronique.

Inversement, un auteur qui souhaite voir apparaître dans la table d'index des références aux sections, aura tendance à **poser des marques sur les sections** et non sur les éléments contenus dans une section qu'il souhaite référencer. Il perdra ainsi l'avantage d'avoir une référence précise dans la section lors de la manipulation du document électronique. En effet, les références dans une table d'index peuvent renvoyer soit au numéro de la page, soit au numéro de la section qui contient l'élément. Il n'y a pas à déplacer les marques pour obtenir des références aux sections. Il suffit de modifier une option et de demander à reconstruire les tables d'index.

Comme les systèmes de PAO utilisent généralement deux marques (Frame-Maker, Word, LaTeX) pour délimiter un long passage, les utilisateurs sont surpris de n'avoir à poser qu'**une paire de marques** pour obtenir un intervalle de pages de références, par exemple.

Enfin, l'absence de commande de déplacement est particulièrement gênante dans le cas des marques d'index: il faut utiliser le couper/coller qui a pour effet de déplacer à la fois les marques et la zone marquée. En conséquence, il faut ensuite coller à nouveau la zone marquée à son emplacement d'origine.

4.2 Vers une indexation plus automatique

L'évolution des méthodes d'indexation automatique incite à penser que la connaissance de la structure du document à indexer offre des éléments supplémentaires qui permettront encore d'affiner les résultats obtenus par les méthodes générales d'indexation automatique [Salton 89a] [Salton 89b]. Comme les travaux de [Kerkouba 85] le montrent, l'importance d'un mot ou d'un groupe de mots dépend aussi de la position de ce mot ou de ce groupe dans le document: s'il est placé dans l'introduction ou dans le résumé, son importance sera plus grande que s'il est dans une annexe ou une figure. Il serait donc intéressant d'intégrer dans Grif un outil d'**indexation automatique** capable de faire des suggestions pertinentes aux auteurs en exploitant la connaissance de la structure des documents.

La commande **Insérer des marques** s'appuie déjà sur la structure du document: elle permet d'insérer des marques dans un document structuré en tenant compte du texte contenu, mais aussi de la position de ce texte dans la structure ou de ses attributs. C'est un premier pas vers l'indexation automatique de documents structurés.

Une première extension de cette commande devrait permettre d'indexer automatiquement dans le document des termes définis dans une liste fournie par l'auteur (et non saisie interactivement, mot à mot). Par exemple, à partir de la table d'index du rapport d'activité de l'année 1991, il sera possible d'extraire une liste de mots-clés. Cette liste sera ensuite utilisée pour indexer automatiquement le rapport d'activité de l'année suivante. La liste des mots-clés peut ensuite être complétée dans le rapport d'activité en cours d'édition.

4.3 Vers une présentation plus homogène des termes

L'utilisation des schémas de structure et de présentation de documents structurés, manipulés par Grif, permet d'assurer que les index produits ont une présentation homogène et une certaine cohérence (contrôle des renvois); par contre, le **développement d'outils spécialisés** est encore nécessaire pour améliorer le contenu des index.

Il faut aider l'utilisateur à remplir le contenu de ces index : aide pour déterminer l'ordre et l'importance des mots, pour construire des entrées multi-niveaux; mais aussi, aide pour harmoniser le format des termes (typographie des noms de personne, ordre des clés et sous-clés), pour regrouper des références rapprochées (en tenant compte de la structure logique du document), etc.

4.4 Conclusion et perspectives

Dans ce rapport, nous avons décrit les index électroniques de Grif qui permettent d'accéder à un document électronique à partir de son contenu. Le document ainsi manipulé est un hypertexte, dans lequel l'utilisateur peut suivre des liens entre tables d'index et contenu du document. Alors que l'auteur d'un document structuré est plutôt guidé par la structure du document, le lecteur peut être guidé

par la structure dynamique des index introduite par l'auteur pour faciliter l'accès au contenu du document.

Plus précis que les index imprimés qui font uniquement référence aux pages d'un document, les index électroniques permettent de retrouver dans le document les éléments, de granularité variable, qui traitent du sujet recherché: un mot, une phrase, des paragraphes ou même des sections entières peuvent être affichés en désignant une entrée de l'index électronique.

Grif offre donc un certain nombre de mécanismes qui permettent de produire des index électroniques. Il serait intéressant d'y apporter encore certaines améliorations qui contribueront au confort d'utilisation des index. Parmi les améliorations envisagées, citons: l'intégration d'un outil d'indexation automatique capable d'indiquer directement à l'auteur du document structuré les meilleurs termes à indexer et sous quelle forme les indexer, la mise à jour dynamique du contenu des tables d'index lors de l'édition des marques et la création d'une table d'index à partir d'une liste de termes préétablie, externe au document.

4.5 Remerciements

Cette étude des index a été réalisée dans le cadre du projet OPÉRA. Je tiens à remercier tous ceux qui y ont participé: Jacques André, Philippe Louarn, Éric Picheral, Vincent Quint et Irène Vatton. La mise en commun de ces compétences a permis que les index électroniques existent maintenant dans Grif. Cette réalisation est la démonstration qu'un projet réparti géographiquement (entre Rennes et Grenoble) peut fonctionner, dès lors que nous disposons de moyens de communications efficaces (merci à l'atelier de l'Irisa). Les réunions de l'équipe, aussi bien à Rennes qu'à Grenoble, les fréquentes discussions à toutes les étapes du projet ont permis à chacun d'y apporter sa contribution: depuis la réflexion initiale (de Jacques), la comparaison avec les systèmes existants (de Philippe, Éric et Hélène), la spécification des index, l'intégration dans Grif (Vincent, Irène et Hélène), jusqu'aux premières expérimentations (Jacques) et à la rédaction de ce rapport qui doit beaucoup à la patiente relecture de Jacques André.

5 Bibliographie

- [André 89] J. André, R. Furuta et V. Quint, éditeurs, *Structured Documents*, Cambridge University Press, 1989.
- [Aurbach 87] R.L. Aurbach, Automated index generation for L^AT_EX, *TUGboat*, 8(2), 201-209, 1987.
- [Bentley 88] J.L. Bentley et B.W. Kernighan, Tools for printing indexes, *Electronic publishing*, 1(1), 3-17, avril 1988.
- [Bruza 90] P.D. Bruza, Hyperindices: A Novel Aid for Searching in Hypermedia, *Hypertext: Concepts, Systems and Applications, Proc. ECHT'90*, A. Risk, N. Streitz et J. André, éditeurs, 109-122, Cambridge University Press, Cambridge, novembre 1990.
- [Chen 88] P. Chen et M.H. Harrison, Index preparation and processing, *Software-practice and experience*, 18(9), 897-915, John Wiley & sons ed., septembre 1988.
- [Dufour 71] M.L. Dufour, *Le Tapuscrit*, CID, Paris, 1971.
- [Framemaker 90] *FrameMaker, Manuel de référence*, Frame Technology Corporation, San Jose, Calif, mai 1990.
- [Frison 91] P. Frison. É. Picheral et H. Richy, *Intégration d'un correcteur orthographique dans l'éditeur structuré Grif*, Rapport de Recherche INRIA-Rennes, n° 1566, décembre 1991.
- [Furuta 88] R. Furuta, V. Quint et J. André, Interactively Editing Structured Documents, *Electronic Publishing - Origination, Dissemination and Design*, 1(1), 19-44, avril 1988.
- [Furuta 90] R. Furuta, Hypertext and electronic publishing, *Hypertext: Concepts, Systems and Applications, Proc. ECHT'90*, A. Risk, N. Streitz et J. André, éditeurs, 347-353, Cambridge University Press, Cambridge, novembre 1990.
- [ISO 1986] *Information processing - Text and office systems - Standard Generalized Markup Language (SGML)*, ISO 8879, octobre 1986.
- [ISO 1989] *Information processing - Text and office systems - Office Document Architecture (ODA)*, ISO 8613, 1989.
- [Kerkouba 85] D. Kerkouba, Indexation automatique et aspects structurels du texte, *RIAO 85*, 227-249, Grenoble, France, mars 1985.
- [Knuth 84] D. Knuth, *The TeXbook*, Addison-Wesley, Reading, Massachusetts, 1984.

- [Lamport 86] L. Lamport, *L^AT_EX: A Document Preparation System*, Addison-Wesley, Reading, Massachusetts, 1986.
- [Lamport 87] L. Lamport, *MakeIndex: an index processor for L^AT_EX*, février 1987.
- [Louarn 90] P. Louarn, Traitement d'index avec L^AT_EX, *Cahiers GUTenberg*, n° 7, novembre 1990.
- [Quint 86] V. Quint et I. Vatton, Grif: an Interactive System for Structured Document Manipulation, éd. J.C. van Vliet, *Text Processing and Document Manipulation*, Cambridge University Press, 200-213, avril 1986.
- [Quint 90] V. Quint, M. Nanard et J. André, Towards Document Engineering, *EP'90*, éd. R. Furuta, Cambridge University Press, septembre 1990, 17-29.
- [Quint 92a] V. Quint et I. Vatton, *Hypertext Aspects of the Grif Structured Editor: Design and Applications*, Rapport de Recherche INRIA-Grenoble, n° 1734, juillet 1992.
- [Quint 92b] V. Quint et I. Vatton, Combining Hypertext and Structured Documents in Grif, à paraître dans *ECHT'92 Proc.*, Milan, décembre 1992.
- [Salton 89a] G. Salton, *Automatic Text Processing*, Addison-Wesley, Reading, Massachusetts, 1989.
- [Salton 89b] G. Salton, *A comparison of book indexing methods*, TR 89-1033, Cornell University, Ithaca (NY), août 1989.
- [Stotts 91] P.D. Scotts et R. Furuta, Hypertext 2000: Databases or documents?, *Electronic Publishing*, 4(2), 119-121, juin 1991.
- [Word 89] Word, *Manuel de référence de Microsoft Word, Word 4 pour Apple Macintosh*, Microsoft Corporation, Ref 505959, 1984-1989.

LISTE DES DERNIERES PUBLICATIONS INTERNES PARUES A L'IRISA

- PI 668 TWO COMPLEMENTARY NOTES ON SKEWED-ASSOCIATIVE CACHES
André SEZNEC
Juillet 1992, 10 pages.
- PI 669 PARALLELISATION D'UN ALGORITHME DE DETECTION DE MOUVEMENT
SUR UNE ARCHITECTURE MIMD
Fabrice HEITZ, Sergui JUFRESA, Etienne MEMIN, Thierry PRIOL
Juillet 1992, 34 pages.
- PI 670 UN RESEAU SYSTOLIQUE INTEGRE POUR LA CORRECTION DE FAUTES
DE FRAPPE
Dominique LAVENIER
Juillet 1992, 120 pages.
- PI 671 EARLY WARNING OF SLIGHT CHANGES IN SYSTEMS AND PLANTS WITH
APPLICATION TO CONDITION BASED MAINTENANCE
Qinghua ZHANG, Michèle BASSEVILLE, Albert BENVENISTE
Juillet 1992, 32 pages.
- PI 672 ORDRES REPRESENTABLES PAR DES TRANSLATIONS DE SEGMENTS DANS
LE PLAN
Vincent BOUCHITTE, Roland JEGOU, JeanXavier RAMPON
Juillet 1992, 8 pages.
- PI 673 AN EXCEPTION HANDLING MECHANISM FOR PARALLEL OBJECT-ORIENTED
PROGRAMMING
Valérie ISSARNY
Août 1992, 36 pages.
- PI 674 A CALCULUS OF GAMMA PROGRAMS
Chris HANKIN, Daniel LE METAYER, David SANDS
Juillet 1992, 32 pages.
- PI 675 EVALUATION DES PERFORMANCES D'UN NOYAU DE SIMULATION
REPARTIE
Philippe INGELS, Carlos MAZIERO
Septembre 1992, 36 pages.
- PI 676 FONT METRICS
Jacques ANDRE
Septembre 1992, 20 pages.
- PI 677 GRIF ET LES INDEX ELECTRONIQUES
Hélène RICHY
Septembre 1992, 40 pages.

ISSN 0249 - 6399